UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. 12

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle) 6

AN ANALOGY PERMITTING MAXIMUM LIKELIHOOD ESTIMATION BY A SIMPLE MODIFICATION OF GENERAL LEAST SQUARES ALGORITHMS.

5. TYPE OF REPORT & PERIOD COVERED 9

TECHNICAL REPORT.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s) 10

Louis D. Homer [redacted] R. Clifton Bailey

8. CONTRACT OR GRANT NUMBER(s)

Interagency Agreement No. 2-Y01-AI-50001 between NMRI and NIH

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Naval Medical Research Institute
Bethesda, Maryland 20014

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

M0098-PN.01.0004
Report No. 1

11. CONTROLLING OFFICE NAME AND ADDRESS

Naval Medical Research & Development Command
Bethesda, Maryland 20014

12. REPORT DATE 11

Sep[redacted] 1977

13. NUMBER OF PAGES

8

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

Bureau of Medicine & Surgery
Department of the Navy
Washington, D.C. 20372

12 11p.

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

63706N

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

16 MΦΦ98 MΦΦ99

17 MΦΦ98PNΦ1, MΦΦ99PNΦ2

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum likelihood, nonlinear least squares, iterative estimation, curve fitting, optimisation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The basis for a method of modifying general nonlinear least squares algorithms for maximum likelihood estimation is developed. The resultant algorithm provides a direct general approach to maximum likelihood estimation. Only the likelihood function for a single observation need be specified. Neither second partial derivatives nor weights are required. First partial derivatives may be obtained numerically and are used to compute the covariance matrix of the estimated parameters.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED 249 650

LB

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AN ANALOGY PERMITTING MAXIMUM LIKELIHOOD ESTIMATION BY A SIMPLE MODIFICATION OF GENERAL LEAST SQUARES ALGORITHMS

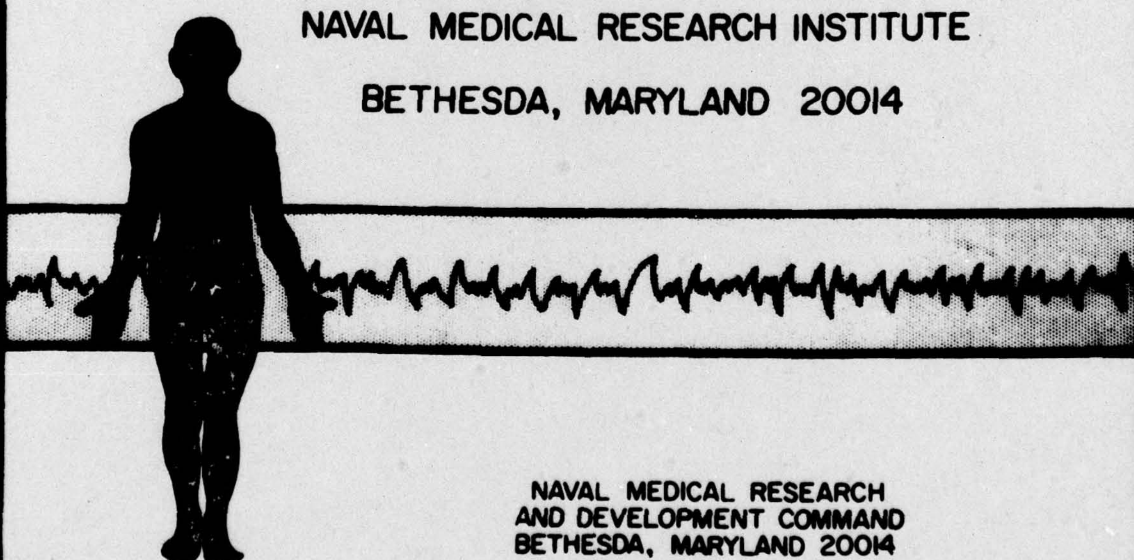
Louis D. Homer, M.D., Ph.D.

and

R. Clifton Bailey, Ph.D.

NAVAL MEDICAL RESEARCH INSTITUTE

BETHESDA, MARYLAND 20014



NAVAL MEDICAL RESEARCH
AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND 20014

Work Unit Nos: M0098-PN.01.0004
M0098-PN.01.0037
M0099-PN.02.5012

AN ANALOGY PERMITTING MAXIMUM LIKELIHOOD ESTIMATION BY A SIMPLE MODIFICATION OF GENERAL LEAST SQUARES ALGORITHMS

Louis D. Homer, M.D., Ph.D.

and

R. Clifton Bailey, Ph.D.

Biometrics Division

Engineering Support Department

Naval Medical Research Institute

Bethesda, Maryland 20014

September 1977

This work was supported in part by the Naval Medical Research and Development Command, Work Units Nos. M0098-PN.01.0037, M0099-PN.02.5012, and M0098-PN-01.0004, and in part by the National Institutes of Health, Interagency Agreement No. 2-Y01-AI-50001. The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the U.S. Navy Department or the naval service at large.

AN ANALOGY PERMITTING MAXIMUM LIKELIHOOD ESTIMATION BY A SIMPLE MODIFICATION OF GENERAL LEAST SQUARES ALGORITHMS

Louis D. Homer and R. Clifton Bailey

Many estimation problems require nonlinear least squares estimation or other iterative maximum likelihood estimation. Whereas many general programs for least squares regression exist, general programs for maximum likelihood estimation are less common.

General discussions of the maximum likelihood method can be found in numerous papers. A recent text by Edwards [5] should serve as a good introduction. Several algorithms for general least squares estimation are conveniently summarized by Draper and Smith [4, Ch. 10]. A more extensive treatment of algorithms for non-linear parameter estimation is given by Bard [2]. Our technique for modification of general least squares algorithms is not a weighted least squares approach as presented by Bradley [3] and implemented by Jennrich and Moore [6]. Since our approach is more direct, we anticipate that comparisons will favor it as a basic computing algorithm. The converted computer program we use requires the user to supply a subroutine with the log likelihood function. Weights are not required. No programming for first or second partial derivatives need be added. The simplicity of conversion is also appealing.

We propose to summarize the major steps required to convert a general least squares program to a general maximum likelihood program.

The typical least squares problem with n observations, a vector $\underline{x}$ of independent variables, and k parameters to be estimated consists in finding parameters $\hat{\underline{\theta}}' = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ to minimize

$$S(\hat{\underline{\theta}}) = \sum_{i=1}^{n} (y_i - f(\hat{\underline{\theta}}, \underline{x}_i))^2 . \tag{1}$$

The usual general program requires the user to specify the form of f in a subroutine or a few lines of coding and to choose starting values $\underline{\theta}_o$. The more versatile programs compute numerical approximations to the partial derivatives of f with respect to the parameters

$$\underset{(n\times k)}{P} = \frac{\partial f}{\partial \underline{\theta}'} \tag{2}$$

and also compute

$$\underset{(n\times 1)}{\underline{g}} = P'(\underline{y} - \underline{f}). \tag{3}$$

A new estimate $\underline{\theta}_1$ is obtained from the relation

$$\underline{\theta}_1 = \underline{\theta}_o + (P'P)^{-1}\, \underline{g}. \tag{4}$$

Next $\underline{\theta}_2$ may be obtained from $\underline{\theta}_1$, and so on until successive changes in $\underline{\theta}$ or S are deemed suitably small and the $\underline{\theta}$ used to compute the smallest S is accepted as $\hat{\underline{\theta}}$.

The iterative maximum likelihood problem consists in finding $\underline{\theta} = \hat{\underline{\theta}}$ to maximize

$$L = \prod_{i=1}^{n} \ell_i(\underline{\theta}, \underline{x}_i), \tag{5}$$

where ℓ_i is the likelihood of the i^{th} observation and L is the

likelihood function. Again we have n observations comprising $\underline{x}$ variables for each observation and k parameters to estimate. We assume the existence of the regularity conditions, like Kendall and Stuart [7]. Maximizing log L is done by obtaining successive $\underline{\theta}$'s by

$$\underset{(k\times 1)}{\underline{\theta}_r} = \underline{\theta}_{r-1} - \underset{(k\times k)}{\left\{\frac{\partial^2 \log L}{\partial\theta_i\,\partial\theta_j}\right\}^{-1}} \underset{(k\times 1)}{\frac{\partial \log L}{\partial \underline{\theta}}} \tag{6}$$

where θ_i is the i^{th} element of the vector of parameters $\underline{\theta}$. Alternatively, one might invert the expected value of the (k x k) matrix above. Neither direct use of the matrix nor its expected value lends itself conveniently to the construction of general programs requiring the user to supply only the form of ℓ_i or log ℓ_i. Additional specifications for first and second partial derivatives of log L or the expectations would have to be made.

If, however, we choose

$$\underline{\theta}_r = \underline{\theta}_{r-1} + \left\{\sum_{i=1}^{n}\left(\frac{\partial \log \ell_i}{\partial \underline{\theta}}\right)\left(\frac{\partial \log \ell_i}{\partial \underline{\theta}}\right)'\right\}^{-1} \sum_{i=1}^{n} \frac{\partial \log \ell_i}{\partial \underline{\theta}}, \tag{7}$$

then letting $\underline{u}$ be a vector of n ones and letting

$$\underset{n\times k}{P} = \frac{\partial \log \underline{\ell}}{\partial \underline{\theta}'} = \{\partial \log \ell_i/\partial\theta_j\} \tag{8}$$

and

$$\underline{g} = P'\underline{u}, = \{\partial \log L/\partial\theta_j\}, \tag{9}$$

we may obtain

$$\underline{\theta}_r = \underline{\theta}_{r-1} + (P'P)^{-1}\ \underline{g}. \tag{10}$$

This is the same form as (4), hence it is possible to write general maximum likelihood programs which are as versatile and easy to use as the general nonlinear least squares programs. Conversion of nonlinear least squares programs for use in maximum likelihood estimation is easy.

If the regularity conditions mentioned before hold, we have

$$-E\left\{\frac{\partial^2 \log L}{\partial\theta_i\ \theta_j}\right\} = E(P'P), \tag{11}$$

so that $(P'P)^{-1}$ is a suitable estimate of the covariance matrix of the estimated parameters. In fact, it is a generalized compromise between using the inverse of $\frac{\partial^2 \log L}{\partial\theta_i\ \theta_j}$ and the inverse of its expectation.

Recall that

$$\frac{\partial^2 \log L}{\partial\theta_i\ \partial\theta_j} = \sum_{m=1}^{n}\left[-\frac{1}{\ell_m^2}\ \frac{\partial \ell_m}{\partial\theta_i}\ \frac{\partial \ell_m}{\partial\theta_j} + \frac{1}{\ell_m}\ \frac{\partial^2 \ell_m}{\partial\theta_i\ \partial\theta_j}\right] \tag{12}$$

and

$$\left(\frac{\partial \log L}{\partial\theta_i}\right)'\left(\frac{\partial \log L}{\partial\theta_j}\right) = \sum_{m=1}^{n}\left[\frac{1}{\ell_m^2}\ \frac{\partial \ell_m}{\partial\theta_i}\ \frac{\partial \ell_m}{\partial\theta_j} + \sum_{\substack{h=1\\h\neq m}}^{n}\frac{1}{\ell_m\ \ell_h}\ \frac{\partial \ell_m}{\partial\theta_i}\ \frac{\partial \ell_h}{\partial\theta_j}\right]. \tag{13}$$

To facilitate comparisons we rewrite 12 and 13:

$$\left\{\frac{\partial^2 \log L}{\partial\theta_i\ \partial\theta_j}\right\} = -P'P + R \tag{14}$$

and

$$\left\{ \left(\frac{\partial \log L}{\partial \theta_i} \right) \left(\frac{\partial \log L}{\partial \theta_j} \right) \right\} = P'P + T \tag{15}$$

The expectation of R is zero. This may be seen by noting that the expectation of T is zero and that the expectation of the left-hand side of 14 is the negative of the expectation of the left-hand side of 15. Since we drop R, our general approach uses a partial expectation, dropping terms known generally to have expectation zero. In the particular case of a multinomial, - P'P is the matrix of second partial derivatives, and consequently the algorithm suggested is exactly equivalent to the algorithm obtained by using the expectation of the second partial derivatives. In the case of estimating the variance and mean of a normal distribution, the corresponding P'P includes off-diagonal elements whose expectation is zero but whose numerical value will only be zero if the third sample moment about the mean happens to be zero. In this case the suggested algorithm differs from that using the expectation of the second partial derivatives because of these possibly nonzero off-diagonal elements. We may regard the operation of taking the expectation of the second partial derivatives of log L as having a general answer which we take advantage of and a particular part which we ignore to preserve the generality and simplicity of our numerical approach. In some instances where more exact results are desired, E(P'P) may be easier to compute than $E \left\{ \frac{\partial^2 \log L}{\partial \theta_i \, \partial \theta_j} \right\}$.

We emphasize that this compromise in the choice of an estimator of the covariance matrix means that one can readily convert least squares programs to maximum likelihood estimation, and these programs need not require the user to supply specific programming for the partial derivatives.

Other changes we suggest are in the accumulation of - log L instead of $S(\underline{\theta})$. We suggest - log L instead of log L since this makes it a minimization problem and many least squares programs check to make sure that $S(\underline{\theta}_{j+1}) < S(\underline{\theta}_j)$ before accepting the new set of parameters. Computation of the covariance matrix in the least squares case is

$$V = \hat{\sigma}^2 (P'P)^{-1}. \tag{16}$$

For the maximum likelihood case the scalar multiplication must be removed, that is,

$$V = (P'P)^{-1}. \tag{17}$$

Other details involving input and output would surely have to be revised but should not pose difficult problems. Since these are apt to be specific to the particular program being converted, an extensive discussion here seems pointless.

A detailed example of conversion is available for a 138-line BASIC program [1] originally designed for nonlinear least squares regression on an interactive time-sharing system using Marquardt's algorithm. The program requires modification of 7 lines for conversion to a general maximum likelihood program. The converted

program requires the user to supply only the programming needed to calculate the log likelihood for an individual observation. This will usually be just a few lines of code. The output includes parameter estimates, the log likelihood, and the covariance matrix. Techniques for using the program output to obtain confidence regions about functions of the parameters by means of propagation of error formulas are also illustrated.

REFERENCES

[1] Bailey, R. C. and Homer, L.D., "Iterative Parameter Estimation," Naval Medical Research Institute, Naval Medical Research and Development Command, Bethesda, Maryland, September 1976.

[2] Bard, Y., *Nonlinear Parameter Estimation,* New York: Academic Press, 1974.

[3] Bradley, E. L., "The Equivalence of Maximum Likelihood and Weighted Least Squares Estimates in the Exponential Family," Journal of the American Statistical Association, 68 (March 1973), 199-200.

[4] Draper, N. R. and Smith, H., *Applied Regression Analysis,* New York: John Wiley and Sons, Inc., 1966, 263-304.

[5] Edwards, A. W. F., *Likelihood,* Cambridge, Mass.: University Press, 1972.

[6] Jennrich, R. I. and Moore, R. H., "Maximum Likelihood Estimation by Means of Nonlinear Least Squares," Technicial Report No. 9, Health Sciences Computing Facility, University of California at Los Angeles, July 1975.

[7] Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics, Vol. 2,* 3rd ed., New York: Hafner Publishing Co., 1974, Ch. 17 & 18.